JRC Technical Notes

On e-testing: an overview of main issues

- Background note -

Ângela Guimarães Pereira, Friedrich Scheuermann

EUR 23057 EN - 2007

JRC
EUROPEAN COMMISSION

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

European Commission
Joint Research Centre

Contact information
Address: Unit G09, Knowledge Assessment Methodologies (KAM)
TP-361, Via Enrico Fermi, 1; 21020 Ispra (VA), Italy
E-mail: angela.pereira@jrc.it, friedrich.scheuermann@jrc.it
Tel.: +39-0332-78.6111
Fax: +39-0332-78.6492

http://www.jrc.ec.europa.eu

JRC 36015

EUR 23057 EN
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

1. Background

The European Council (2006/C172/01) conclusions[1] on the European Indicator of Language Competence ask for measures for objective testing of skills in first and second foreign languages based on the Common European Framework of Reference for Languages (CEFR). The Council invites the Commission to assist Member States (MS) to define the organisational and resource implications for them of the construction and administration of tests, including looking into the possibility of adopting e-testing as the means to administer the tests. Electronic testing could improve the effectiveness, i.e. improve identification of skills, and efficiency, by reducing costs (financial efforts, human resources etc.).

A variety of piloting activities have been undertaken to support recent discussions about integrating information technologies (IT) into the tasks of assessing skills. These activities aim at verifying the strengths and weaknesses, potentials and barriers posed in terms of their application in real situations. Such experiences are not sufficiently documented yet but results so far suggest that IT based tools could support the assessment process and the analysis of results. In terms of language skills assessment there are certain specific barriers to be taken into account, and that will be discussed in the reminder of the document.

This document is organised in the following manner; first a working definition of computer based testing is presented, followed by some information regarding state of the art and the use of computer-based testing (CBT). Section two outlines the main advantages and challenges of computer based assessments, with a focus on the field of language competence. The final, section is dedicated to a rough discussion of costs.

2. General issues about computer based testing

2.1 Working definition

Various terms are used to describe the use of a computer for assessment purposes. These include:

1. Computer-Assisted Assessment or Computer-Aided Assessment (CAA)
2. Computer-Mediated Assessment (CMA)
3. Computer-Based Assessment (CBA)
4. Online Assessment or Internet-Based Assessment (IBT)
5. E-Assessment
6. Computer Adaptive Testing (CAT)

Although these terms are commonly used interchangeably, they have distinct meanings (please see the Appendix where various definitions are explained). In all cases computer technology is supportive of the tasks related to assessment[2]. This document adopts the expression of computer-based testing (or *e-testing)* since the delivery of tests in electronic format is designated as such in the European Council conclusions.

[1] Council Conclusions on the European Indicator of Language Competence. OJ C172, 25/07/2006. Page 1.

[2] In addition to this, the word “assessment” is often replaced by the word “testing”, which adds on the variety of meanings.

2.2 Evolution of deployment: cases of language testing

There is some evidence that e-testing is no longer viewed as a mere means for automation of an existing process, but indeed reconceptualises it with desirable improvements and inevitable challenges. CBT is being used in various assessment contexts, including placement testing, certification, *formative* and *summative*[3] evaluation.

In the USA, the development from paper-and-pencil testing to computer based options, using computer adaptive testing (CAT) truly began when in 1994 the US National Council of State Boards of Nursing launched a e-testing version of its licensing exam (NCLEX/ CAT) which was followed by a CAT version of the Graduate Record Examination (GRE), the test which is used as an entry exam for graduate (post-bachelor) university studies[4]. Since then, a number of testing activities has witnessed a substantial growth worldwide. Below, we briefly review initiatives in the domain of language testing of some relevance in Europe.

In the language domain, the most important products and services are the ones developed by the American "Educational Testing Service" (ETS) which internationally delivers large-scale testing services, such as SAT and TOEFL. TOEFL is nowadays offered in all European countries only as Internet-Based Testing mode. TOEFL tests comprise 3 sections: Listening Comprehension; Structure and Written Expression and Reading.

Apart from this example from the Non-Profit sector, many tools and applications are being developed by commercial enterprises with specific services on well-focussed areas (see Appendix). Commercial methodologies, underlying organisational concepts and codes/algorithms usually are unpublished and cannot be adopted in other contexts.

Whereas the use of computers in the assessment of skills is increasingly common in the US (http://www.jmu.edu/icba/prodserv/adaptexoverview.htm), there are not too many experiences at European level; small scaled experiences are reported at a project level, e.g.:

- DIALANG: The DIALANG[5] project is about computer based language testing. It is an assessment system intended for language learners who want to obtain diagnostic information about their language proficiency, providing also advice about how to improve language skills. DIALANG is Internet based freeware, currently managing diagnostic tests in 14 different European languages.
- TAO[6]: TAO was developed by the EMACS research unit of the Univ. of Luxembourg and the CITI department of the Centre de Recherche Public Henri Tudor, standing for *Test Assisté par Ordinateur*. Recently it has been used for a pre-test suite of French Knowledge (TCF) in the Luxembourg with 248 test takers. The idea is to have all students in Luxembourg to be tested for French Knowledge using this system in 2007 (see Appendix).
- PISA pilot studies: (additional national pilots 2003, CBAS 2006)

[3] 'Summative evaluation' refers to the observation of states (knowledge, skills, competence) at a **fixed time**; 'formative evaluation' to processes (teaching, learning etc.) for a **period** of time.
[4] Mills, C. N and M. Steffen (2000). "The GRE Computer Adaptive Test: Operational Issues", in Van den Linden and Glass (2000), pp. 75-100.
[5] See for instance Anderson, J. C. 2005: Diagnosing Foreign Language Proficiency. The Interface between Learning and Assessment. London: Continuum. And http://www.dialang.org.
[6] See section 6 and Appendix for further information. Also: http://www.tao.lu.

3. Promises and Challenges of Computer Based Testing

The advantages and challenges of CBT tools compared with traditional pencil and paper testing can be presented in different ways. In this document we choose to present them from the point of view of the most relevant "social actors" involved in the process of testing: test takers, test administrators and test developers. It must be noted that the challenges and promises have to be described against an assessment context; the types of advantages that e-testing may have for *summative* testing are different from those of large scale surveys of skills assessments. Moreover, the challenges in this field are different if one deals with mathematic skills or language skills. Hence, where appropriate we will relate challenges and advantages to the language skills assessment context.

The nature of challenges and advantages cover many dimensions: economic, technical, pedagogical, psychological, organisational, etc. These will be reflected on the typology described below.

3.1 The **test taker** perspective

The test taker perspective is reflected in the following aspects: scheduling of tests, speed of test execution, accessibility, availability, skills requirements, active learning and other types of motivation and engagement. These will be described in detail below.

- **Scheduling of tests**: test takers may decide the times at which tests are taken in some assessment contexts.
    - Advantages: tests may be offered on demand, at times convenient for tests takers. N.B. in high stakes assessments, tests may have to be carried out under some supervision and therefore in dedicated premises and at specific times;

- **Speed of test execution** (and administration): this relates to the speed at which the test taker performs the test but also to the aspect of self-pacing.
    - Advantages: especially with CAT, tests based on computers are said to be faster than traditional ones because test takers respond to items closer to their actual skills. In some testing contexts self-pacing is a relevant advantage.

- **Accessibility:** to tests by test takers with disabilities.
    - Advantages: possibility of using adaptive devices is improved (e.g. screen readers, Braille displays, screen magnification, self-voicing we browsers, etc.); visual aids go beyond text.

- **Testing availability:** Internet based tests can improve test-takers access regardless of geography
    - Advantages: many related to mobility; oftentimes no need to install applications in the computer of the test-taker.
    - Challenges: possibility of system failure;

- **Active learning:** many e-testing software applications provide algorithms for automatic instant scoring which can be fed back to the test taker while carrying out the test.
    - Advantages: this process is rewarding and reassuring for the test taker.

- **Skills:** additional skills are required for test takers, test authors and test administrators;
    - Challenges*:* the skills required to carry out tests in a computer are IT related: typing, mouse navigation, key combinations, etc.: training may be required; reading from a screen is more fatiguing, especially long passages with scrolling; some test takers may not work with computers and so "thinking on the monitor" may be difficult.

- **Motivation and engagement:** in the testing beyond the technology.
    - Advantages: instant results and immediate diagnostic feedback indicating the candidate's strengths and areas for improvement; CAT technologies have been found to improve test-taking motivation.
    - Challenges: although there is no evidence that anxiety or performance[7] are affected by these types of tests, further research may be needed.

3.2 The **test administrator** perspective

While computerised tests are not intrinsically better than paper-and-pencil tests, there are some distinct advantages of the testing administration process (Parshall et al, 2002)[8]. The test administrator perspective is reflected in the following aspects: monitoring and reporting the test suite, logistics, distribution, interoperability, test administration duration, distribution, security, software interoperability and hardware related matters. These will be described in more detail below.

- **Monitoring and reporting:** many e-testing software applications provide algorithms for processing scores and statistical packages for analysis in a coherent environment
    - Advantages: analytical tasks can be made easier, since data is stored in digital format and can be processed with analytical tool; hence, preparing reports are less time consuming
    - Challenges: scoring algorithms for some skills are still major subjects of current research.

- **Logistics:** implementation of tests
    - Advantages: elimination of complex logistics problems distribution, storage and tracking of test papers;
    - Challenges: other types of logistics arise, including operational issues related to the use of technology; schools/institutions may lack technical expertise that e-testing requires; the infrastructure required (computers, telecommunications, etc.) is not a minor issue.

[7] See for instance, Cassady, J. C. & Gridley, B. E. 2005: The Effects of online formative and summative assessment on test anxiety and performance. In *The Journal of Technology, Learning, and Assessment.* Vol 4 (1). Available from http://www.jtla.org.

[8] Parshall, C. G.; Spray, J. A.; Kalohn, J. C. & Davey, T. 2002: Practical considerations in computer-based testing. Berlin: Springer.

- **Overall duration of test administration**
    - Advantages: time is saved in the scoring operations, as well as if the software is linked to specific analytical packages, analytical work is also more efficient.

- **Distribution:** if Internet based testing: it may be possible to distribute tasks among remote "social actors" i.e. to operate a distributive testing administration and preparation
    - Advantages: distribution of tasks among remote "social actors" who can be geographically remote facilitates co-operation and may save time and costs
    - Challenges: these are related to security; transactions in the Internet may be not so secure; issues with authentication of the personnel involved should also be taken into account.

- **Interoperability:** the possibility of interfacing the CBA software with other analytical applications
    - Advantages: many applications can interface with existing applications used for the analysis of data resulting from tests; the speed of processing is of course much higher.
    - Challenges: compliance with existing technology standards

- **Security:** of test administration and all data transactions need in whole process of testing
    - Advantages: compared with the papers stacks that have to be distributed to issue tests in large scale assessments, a computer based option appears to be more secure
    - Challenges: high stakes assessments can imply issues of security; for instance, platforms that use mySQL databases which are not secure. Hence, this might be an issue in certain forms of assessment.

- **Hardware performance**: have adequate infrastructure to execute tests; ensure tests specific computer resources requirements are met.
    - Challenge: poor performance of hardware may influence the test taker's performance

3.3 The **test developer** perspective

For the test developer, it is important to mention the type of testing that is relevant for skills assessment: CAT. Other aspects that are to be considered form the perspective of a test developer are: measurement algorithms, reusability and adaptability of test content, creative possibilities for item types and also item content.

Computer Adaptive Testing[9]: when dealing with evaluation of skills, an obvious advantage of computer based assessments over the traditional pencil and paper is computer adaptive testing (CAT). CAT is often developed on the basis of item response theory (IRT)[10] psychometric model family. An advantage of this approach is the reduction of test execution time, the test taker being always faced with a "realistic" challenge. Moreover, it is claimed as more secure, since each test taker is given a tailored test, and cheating becomes difficult. CAT is completely dependent on technology and as yet it cannot handle open-ended questions. The pool of items has to be quite large.

- **Measurement algorithms:** the types of assessments that can be carried out (i.e. beyond sequential testing: adaptive testing and others)
    - Advantages: e.g adaptive testing adjusts the difficulty of test items to the ability of the test-taker; in the field of language testing, (adaptive) e-testing is more reliable compared to conventional tests[11]. Hence, it is said to offer better **measurement precision**
    - Challenges: a large item bank has to exist and this requires more effort on the construction of the item bank

- **Reusability and adaptability of test content:** Test questions can be created in 'item banks' and delivered at random.
    - Advantages: item banks can be easily populated and updated; also in many e-testing software, tests can be created by rearranging items, amongst others, cutting out 'battery' testing, i.e. the need to test all candidates at the same time on the same day; it saves time, especially for tests organisation.
    - Challenges: in order to develop a CAT, a large sample of test items needs to be calibrated through pilot testing before actual administration

- **Creativity on item types:** e-testing offers different response options going beyond multiple-choice solutions; e.g. drag and drop answering modes and other types of interactivity.
    - Advantages: This has direct implications on accessibility (also by using drives other than the mouse, other sorts of navigation are possible); allows for construction of new types of test items.
    - Challenges: Handling of open-ended questions is a big limitation; e-testing does not improve the current status; algorithms to process essays are still in their infancy, as well as assessment of "speech". Interactivity offered by computers seems to be the key to find "proxes" for current limits of scoring (e.g. of open-ended questions). More time consuming for those who develop the tests; training and additional skills are required. Moreover, for some disabilities further research on computer navigation systems is needed. Test developers may need specific authoring skills in order to enhancement of items' accessibility.

[9] See for instance: Rudner, L. M. 1998: An on-line interative computer adaptive testing tutorial. Available at http://edres.org/scripts/cat, last accessed: 15/03/2006.
Wainer, H. 1990: Computer Adaptive Testing: a Primer. Hillsdale, N.J.: Lawrence Erlbaum.

[10] See for instance: Baker, F. B. 2001: The Basiics of Item response Theory. Madison, W: ERIC Clearinghouse on Assessment and Evaluation.

[11] See Madsen, H. S. 1991: Computer Adaptive Testing of Listening and Reading Comprehension: The Brigham Young University Approach.", *Computer Assisted Language Learning and Testing: Research Issues and Practice*, edited by P. Dunkel. New York, NY: Newbury House.

- **Item content:** inclusion of multi-media content such as audio, video, animations, rich graphics, etc.
    - o Advantages: innovative assignments may be devised, including more faithful assessment of skills; accessibility is also enhanced, extending the boundaries of paper based assessments
    - o Challenges: more time consuming for those who develop the tests; training and additional skills may be required.

3.4 Language skills assessment challenges

The Council conclusions suggested assessing competence in the 4 receptive and productive skills, but "for practical reasons" to focus on the following areas in the first round: listening comprehension, reading comprehension, writing, whereas testing of speaking skills is left for a later stage. This is presumably due to the fact that currently, productive skills are more difficult to assess:

1) **Listening** comprehension skills (receptive): can easily be implemented and delivered via e-testing.
2) **Reading** comprehension skills (receptive): can easily be implemented and delivered via e-testing.
3) **Writing** skills (productive): at present, skills such as writing of essays would not benefit a great deal if assessed via CBT since there are severe limitations of current assessment algorithms for essays. The only existing robust product in the market is property of ETS (e.g. CriterionSM), consisting of an algorithm for automated and immediate feedback on essay-writing performance (holistic score and annotated diagnostic feedback); at present there is no (open) source code available which could support the integration of these features in other platforms. "Automatic computer-based marking of subjective, free-text responses still operates at basic levels of character or rule recognition. For this reason, the future of subjective testing will depend on human marking, albeit online marking or expensive researching and piloting of more advanced, essay-marking software." [Liam Wynne from Pearson Vue]
4) **Speaking** skills (productive): A more complex task lies on assessing speaking skills. Here specific requirements are given to the IT resources at the user place (allowing speech recognition in terms of hard- and software) as well as the required bandwidth, which is extraordinary high.

Whether undertaken in an electronic mode or not, the most important challenge of assessing productive skills is, in both cases, that heavy investments are needed to deliver and generate results at large-scale level. As demonstrated by PISA, the provision of open questions is rather cost-intensive. Due to the further overall benefits which can be achieved by CBT this should not provoke a general debate on whether CBT is needed or not. However, the amount of human effort and costs are in direct relation to task design and needs to be carefully thought about.

It is interesting to note that research carried out at the University of Luxembourg for testing French language skills according to the French TCF-test system in 4 schools with 248 participants suggests that closure tests ("c-tests") for measuring a global language competence have very similar results to TCF and constitute a cost-effective alternative. However, although some European research is already supporting this approach [Grotjahn 2006[12], Reichert *et al.* 2006[13]], this still needs to be further validated; C-Tests are increasingly used by Higher Education language centres for skills/course identification.

[12] Grotjahn, R. (Ed.) 2006. Der C-Test: Theorie, Empirie, Anwendungen /The C-Test: Theory, Empirical Research, Applications

4. Overview of e-testing costs

Can CBT reduce costs compared to pencil and paper tests? There is no clear answer to be given at this stage, but the increasing uptake of CBT for assessing skills at an international level is an indication for the fact that more effectiveness and/or efficiency is to be achieved with CBT at a general level.

Costs are related to the infrastructure needed for testing (IT, bandwidth etc.), training of testing community (authors, markers etc.), piloting for ensuring valid and reliable results, development and maintenance and administrative issues.

The costs depend on the purpose, the number of countries involved, population to be tested, specification of what type of skills should be tested, requirements/ limitations to be taken into account such as relating to IT infrastructure and on the parameters for testing (e.g. how many questions/tasks, types of questions/tasks). Even when the testing context is specified, savings of applying CBT can hardly be determined since there are no data available for any other type of assessment based on this given scenario. "*As a general rule, online assessment becomes more cost effective the higher the volumes are, while paper-methods are more cost effective for small scale assessments or piloting work." (Dave Bartram, e-mail interview).*

Furthermore, it has been reported that CBT is more cost-effective than traditional P&P solutions, especially in case of objective, closed-ended questions. Computer-supported marking reduces efforts of human involvement as well as logistic (production, transport etc.) and administrative costs are estimated to be lower than in conventional settings. However, "*as attempts are made to transform either objective or subjective paper-based tests into computer-based tests that are more engaging for the candidate, costs increase. Such costs relate to the use of rich media; engaging response mechanisms, and other more advanced technology. Costs increase even further when* [e-testing] *is used not only to make a more engaging test but it is also used to enhance it from a pedagogic point of view. For example, presenting candidates with a task simulation in order to provide a valid assessment of a practical situation will involve the software costs of creating a simulation" (Wynne 2006)*[14]*.*

A combination of paper and pencil and CBT modes would be little cost-effective as indicated by Poggio et al. (2005)[15]; yet further research is needed for detailed analysis. Moreover, existing studies show that modes influence test responses e.g. in terms of expression and structure what might affect the comparability to be achieved (Goldberg et al. 2003[16], Ferris 2002[17]).

For the identification of cost-effective solutions for the assessment of language skills it is important to reflect on the **type of language skills** to be assessed and the item types to be applied. The assessment of *productive* skills will considerably increase costs with each additional open question/task selected due to *human involvement* needed for marking/grading. Assessment alternatives, such as the c-test for global language

---

[13] Reichert, M.; Keller, U. and Martin, R. 2006. Le Test de Connaissance du Français et le C-Test: étude sur la comparabilité entre les deux instruments. Uni. of Luxembourg.
[14] Wynne, L. (v.1), 2006. e-Assessment and value, to be published by Pearson Vue.
[15] Poggio, J.; Glasnapp, D. R.; Xiangdong Yang; 2005. A comparative Evaluation of Score Results from Computerized and Paper & Pencil Mathematics Testing in a large Scale Assessment Program. In: JTLA, Journal of Technology, Learning, and Assessment, Volume 3, Number 6, February 2005. Also available at: www.bc.edu/research/intasc/jtla/journal/pdf/v3n6_jtla.pdf, last accessed: 30.11.2006.
[16] Goldberg, A.; Rusell, M. & Cook, A. 2003. The effect of Computers on Student Writing: A Meta-analysis of Studies from 1992 to 2002. In: JTLA, Journal of Technology, Learning, and Assessment, Volume 2, Number 1, February 2003. Also available at: http://www.bc.edu/research/intasc/jtla/journal/pdf/v2n1_jtla.pdf, last accessed: 30.6.2006.
[17] Ferris, S. 2002. The Effects of Computers on Traditional Writing. In: The Journal of Electronic Publishing, August, 2002, Volume 8, Issue 1, also available at: http://www.press.umich.edu/jep/08-01/ferris.html, last acessed: 30.11.2006

competence assessment (see previous chapter), might offer a very cost-effective solution, but further research is needed.

Some figures might help to get a picture on possible costs: For PISA 2009 the total estimated costs for the implementation of "reading of electronic texts" into the PISA Reading Literacy framework based on existing school computer equipment were calculated to a total of 1.440.000 Euro for a 4 years period (EDU/PISA/GB(2006)31[18], EDU/PISA/GB(2006)27[19]). The calculation is based on the assumption that it will be possible to use the TAO assessment system[20], which is already developed by the University of Luxembourg and that no further investments in platform development are needed. The University of Luxemburg has a laboratory consisting of several servers that manage different activities (modules) regarding testing, the overall investment was around 100 000 EUR, including 8 test delivery servers. With this infrastructure it is estimated that 450 tests per day per delivery server can be delivered.

5. Final reflection

Any of the delivery modes, whether Paper-Pencil and/or computer-based, comprises advantages and challenges which can hardly be compared, especially in relation to estimated costs. As mentioned above the use of CBT includes additional benefits which can be achieved from a organisational, psychological, analytical and pedagogical perspective. Many experts agree on the overall added value and advantages of e-testing in large scale assessments[21]. Furthermore, as already pointed out by research presented to PISA Governing Board, October 2006[22], change of cultural habits e.g. in terms of reading from computers vs. printed material might suggest an on-going change of assessment forms as well.

Overall, CBT is a logic follow-up in the sequence of improvements to be achieved in terms of assessment methodologies, test development, delivery and valorisation of results for multi-purposes.

CBT is a promising option, but it should be carefully examined against the context of language testing. This overview suggests that the adoption of CBT for the assessment of certain language skills is feasible. However, assessing productive language skills (writing, speaking) requires further (human and therefore financial) efforts, which will increase the overall costs to a large extent.

[18] EDU/PISA/GB(2006)31. International option for the assessment of reading of electronic texts. Minutes.
[19] EDU/PISA/GB(2006)27. Assessing the reading of electronic texts / Proposal for inclusion in PISA 2009
[20] See box above and Http://www.tao.lu.
[21] From e-mail interviews carried out in the period of 10-20 November 2006 with some experts in the testing field.
[22] EDU/PISA/GB(2006)31 EDU/PISA/GB(2006)31. International option for the assessment of reading of electronic texts. Minutes.

Appendix

Annex 1: Definitions of computer based assessment activities

- **Computer assisted/mediated assessment** refers to any application of computers within the assessment process; the role of the computer may be extrinsic or intrinsic. It is, therefore, a synonym for e-assessment which also describes a wide range of computer-related activities. Within this definition the computer often plays no part in the actual assessment of responses but merely facilitates the capture and transfer of responses between candidate and human assessor.
- **Computer-based assessment** refers to assessment which is built around the intrinsic use of a computer. This can relate to assessment of IT practical skills or more commonly the on-screen presentation of knowledge tests. The defining factor is that the computer is marking or assessing the responses from candidates.
- **Online assessment/Internet-based assessment** refers to assessment activity which requires the use of the Internet. The Internet is used either in terms of a tool for delivery (e.g. upload, download, administration) or for concrete assessment activities (such as interviews), sometimes there is a combination of both.
- **Computer-based testing / E-testing** describes one potential source of (diagnostic) assessment input. It also refers to the conceptualisation and administration of tests using computers with software applications either commercial or open source.
- **Computer adaptive testing** is an automated form of assessment in which all test takers start at a common starting question, but the score of individual responses then determines the questions the person gets next. This approach to testing is designed to ensure the test taker responds to questions that are close to their true ability level.

Annex 2: Software Applications for testing

There exist a large number of electronic tools on the market[23] supporting assessment activities. Such tools are offered either as specific features of educational platforms that enable the management of multiple-choice items together with the administration and server- or web-based delivery of tests (e.g. Moodle[24]), as survey development tools (e.g. Hot Potatoes[25], WebQuiz[26], Questiontools[27]), as assessment tools (e.g. OpenSurveyPilot[28], CQuest[29], FastTestPro[30], InQsit[31], Perception[32], CASTLE (Computer Assisted Teaching and LEarning)[33], Interactive Question Server[34]) or assessment services (e.g. Pan Testing[35],

[23] See for instance: Plichart, P., Jadoul, R., Vandenabeele, L. and Latour, T. 2004. TAO, a Collaborative distributed computer-based assessment framework built on Semantic Web standards" In International Conference on Advances in Intelligent Systems – Theory and Applications, AISTA 2004 in cooperation with IEEE computer society, 15 – 18 November 2004, Luxembourg.
[24] An open-source Learning Management System including an assessment tool (http://moodle.org/)
[25] Half Baked Software Inc. (http://hotpot.uvic.ca/
[26] SmartLite Software (www.smartlite.it)
[27] Abas UK ltd. (http://www.questiontools.com)
[28] Open-source project http://www.opensurveypilot.org/)
[29] Cogent Computing Corp. (http://cogentcorp.com)
[30] Assessment System Corp. (http://www.assess.com/FastTEST.htm)
[31] Ball State University (http://www.bsu.edu/inqsit/)
[32] QuestionMark Computing Ltd. (http://www.questionmark.com/us/perception/index.htm
[33] University of Leicester (http://www.le.ac.uk/cc/ltg/castle/)
[34] WebMCQ ltd. (http://www.mcqi.com.au/mcqi/shwaa2/eim2/website/products/iqs/intro/index.html)

Assessment Solutions[36], Adaptex Assessment System[37]) covering a wide range of (tailor-made or standard) activities proposed depending on specific needs. Such services are usually offered by commercial enterprises (ASP).

Other commercial and freeware (and indeed open source) tools for Computer-Based Assessment can also be identified via the Internet[38]. Many of them are proprietary, closed, centralised and expensive.

The TAO (*Test Assisté par Ordinateur*) system (Plichart *et al.* 2004 and http://www.tao.lu)[39] is a modular platform for internet-based computer aided testing. The platform allows the management of knowledge pertaining to subjects (individuals whose competencies and knowledge may be assessed), groups of subjects, tests and items (elements of tests requiring an answer from the user). TAO is said to be a flexible and distributed system since it uses meta-data for resource description formalised through Semantic Web standard language RDF/S. In the words of the TAO authors any sort of testing in several domains, including accreditation and even surveying could usefully deploy this open source (OSS).platform. This system is still under development, although a full prototype already exists. The TAO system has not undergone major testing. Also, according to the authors it has much more potential than existing assessment platforms, being a dedicated assessment platform, the elements and properties of which, provide the link with psychometric theory (item parameters and characteristics, testing algorithms etc.) being explicitly built into TAO, but still open for relevant tailoring. The platform is in principle interoperable with other electronic applications.

Its main assets, regard the open shell concept that allows easily specific functionality to be added as a plug-in; currently it includes a variety of assessment models, as well as possibilities for having construction of items other than just multiple choice, in addition to a user friendly interface from the point of view of the test taker. However, the platform is not yet developed on industrial standards due to lack of funding. One of the reasons to go Open Source is to try to boost through a community of users its further developments.

[35] Performance Assessment Network (pan), http://www.pantesting.com
[36] Management Development Systems, http://www.assessmentsolutions.com/
[37] James Madison University (http://cars.jmu.edu/adaptex/)
[38] see JRC report on assessment software, to be published Spring 2008.
[39] Plichart, P., Jadoul, R., Vandenabeele, L. and Latoru, T. 2004. TAO, a Collaborative distributed computer-based assessment framework built on Semantic Web standards" In International Conference on Advances in Intelligent Systems – Theory and Applications, AISTA 2004 in cooperation with IEEE computer society, 15 – 18 November 2004, Luxembourg.

European Commission

EUR 23057 EN – Joint Research Centre
Title: On e-testing: an overview of main issues
Author(s): Ângela Guimarães Pereira, Friedrich Scheuermann
Luxembourg: Office for Official Publications of the European Communities
2007 – 13 pp.
EUR – Scientific and Technical Research series – ISSN 1018-5593

Abstract
The European Council (2006/C172/01) conclusions1 on the European Indicator of Language Competence ask for measures for objective testing of skills in first and second foreign languages based on the Common European Framework of Reference for Languages (CEFR). The Council invites the Commission to assist Member States (MS) to define the organisational and resource implications for them of the construction and administration of tests, including looking into the possibility of adopting e-testing as the means to administer the tests. Electronic testing could improve the effectiveness, i.e. improve identification of skills, and efficiency, by reducing costs (financial efforts, human resources etc.).
A variety of piloting activities have been undertaken to support recent discussions about integrating information technologies (IT) into the tasks of assessing skills. These activities aim at verifying the strengths and weaknesses, potentials and barriers posed in terms of their application in real situations. Such experiences are not sufficiently documented yet but results so far suggest that IT based tools could support the assessment process and the analysis of results. In terms of language skills assessment there are certain specific barriers to be taken into account, and that will be discussed in the reminder of the document.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.